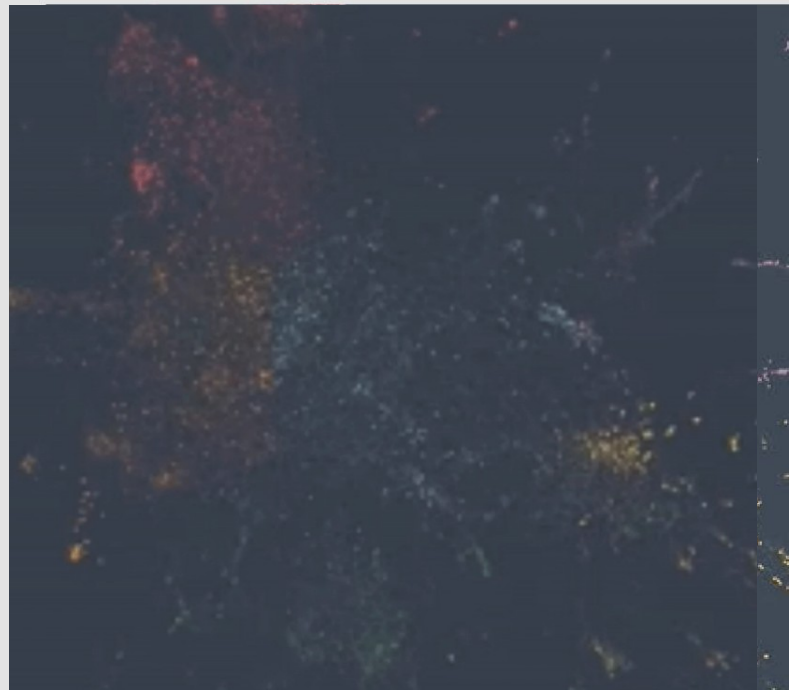


Visualizing meaning and participation at scale from Decidim data, with cosmos.gl

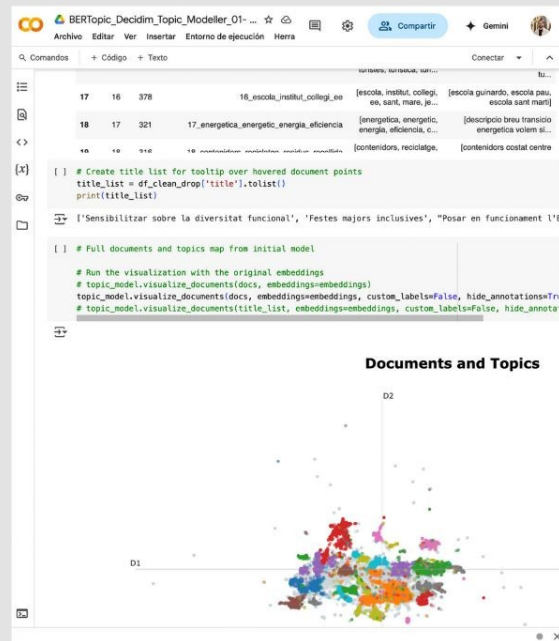
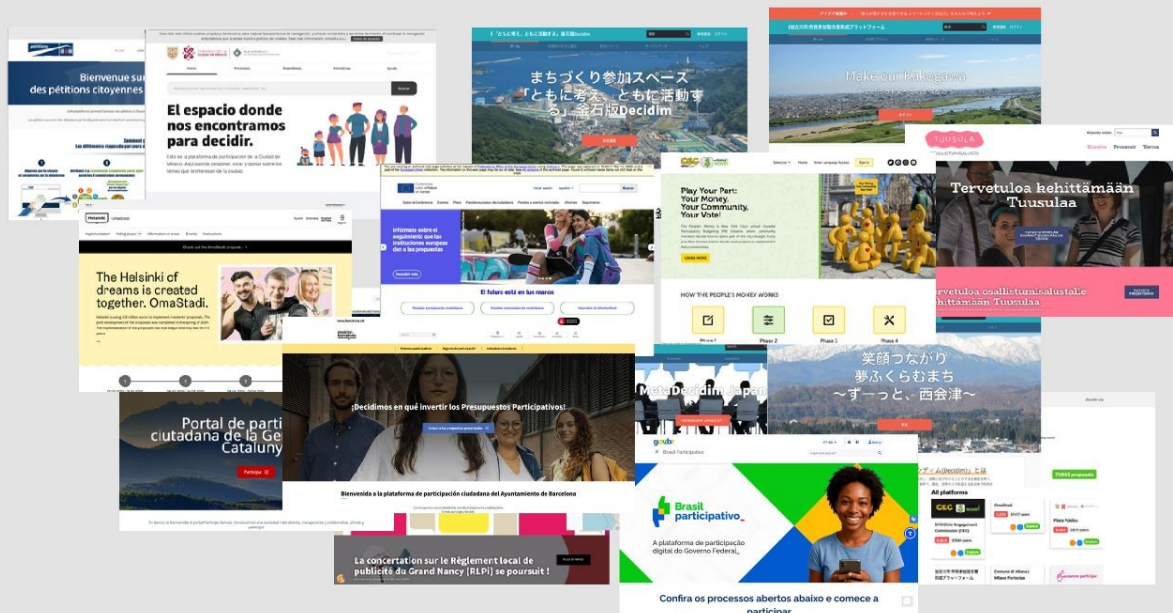
DecidimFest 2025

Lightning talks - Round 2: Artificial Intelligence

Diego Arredondo Ortiz



Making sense of textual data from Decidim-based participatory platforms



Too much text?



Transcriptions and chat
text **from virtual sessions.**



API data from proposals
and comments in
participatory platforms.



Transcriptions and
collective notes from
**deliberative sessions
and assemblies**

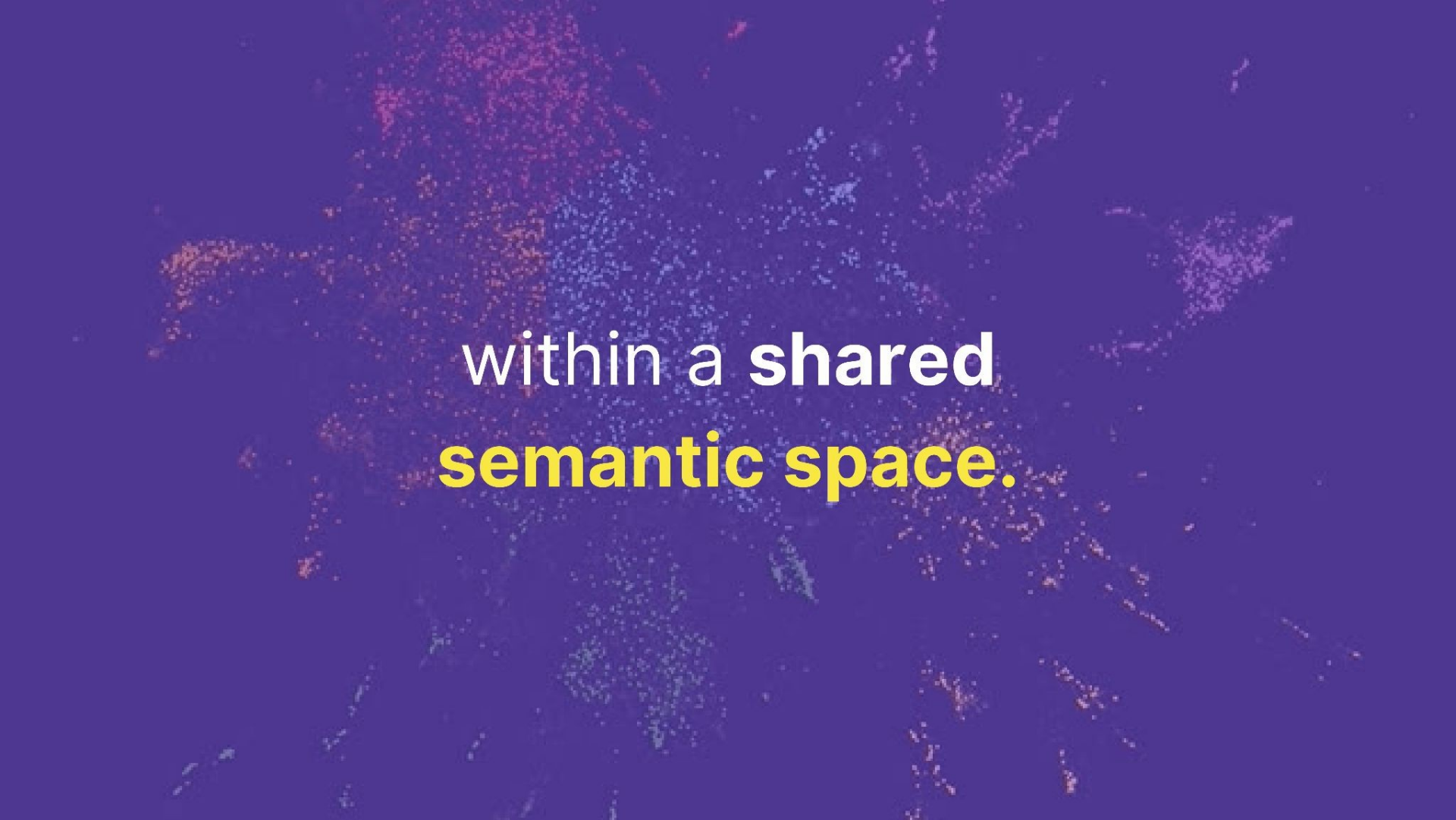
Or better context?



**External information
sources** (laws, relevant
documents, open data, etc.)



Digital and digitized
documents and notes.



within a **shared**
semantic space.

Hover over the points to read proposal contents..

escola - institut - col·legi

dones - sexual - genere

joves - educacio - escola

esport - espai - projecte

salut - persones - atencio

barri - participacio - entitats

guardia - soroll - vigilancia

cultural - culturals - musica

habitatge - habitatges - lloguer

gossos - animals -

bus - autobusos - transport

energetica - energetic - energia

jardins - parc - arbres

bici - bicicletes - carril

LLMs for **retrieval** and **representation** tasks



Unstructured and semistructured data
(text + metadata, audio and image transcriptions).



Tokenization
with open pretrained models



Embeddings
generation



Dimensionality reduction (UMAP), Clustering, topic modelling (BERTopic)

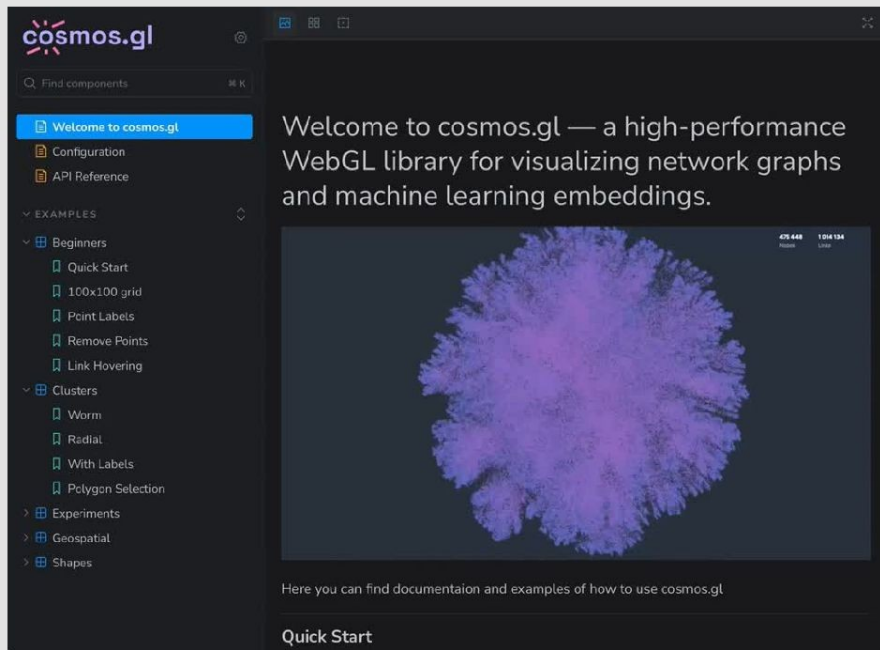


Custom-trained and purposeful categorization



Semantic search for grounded and **intuitive retrieval** (RAG)

Cosmos.gl: a powerful engine for visual exploration of massive and complex datasets (embeddings/graphs)

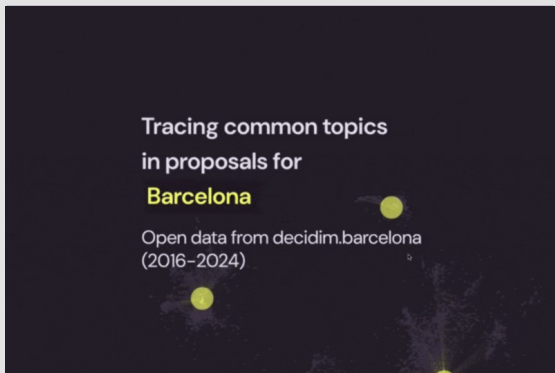


- **Most efficient GPU-accelerated** graphics force layout engine currently available as open source library.
- **Free software license (MIT)**, created by Nikita Rokotyan and currently incubated by OpenJS Foundation.
- **(Growing) documentation** of API reference and examples emerging from different domains, for visualizing large datasets and their relationships (2D embeddings, clusters, networks, labels, images nodes, etc.).

Visualizing shared meaning and interactions at scale with cosmos.gl

Cosmos.gl

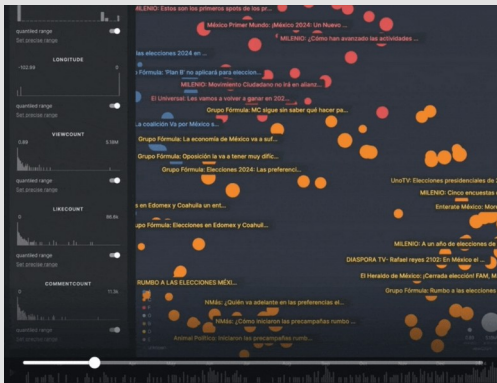
GPU-accelerated force layout library and engine (TypeScript)



MIT License

Cosmograph

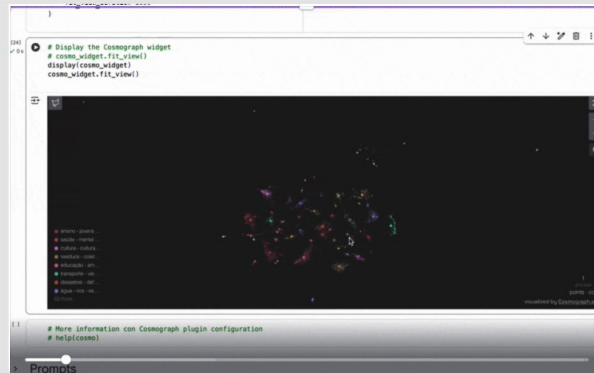
Free(mium) online tool for visualizing embeddings and network graphs



CC BY-NC 4.0 License

Cosmograph Widget

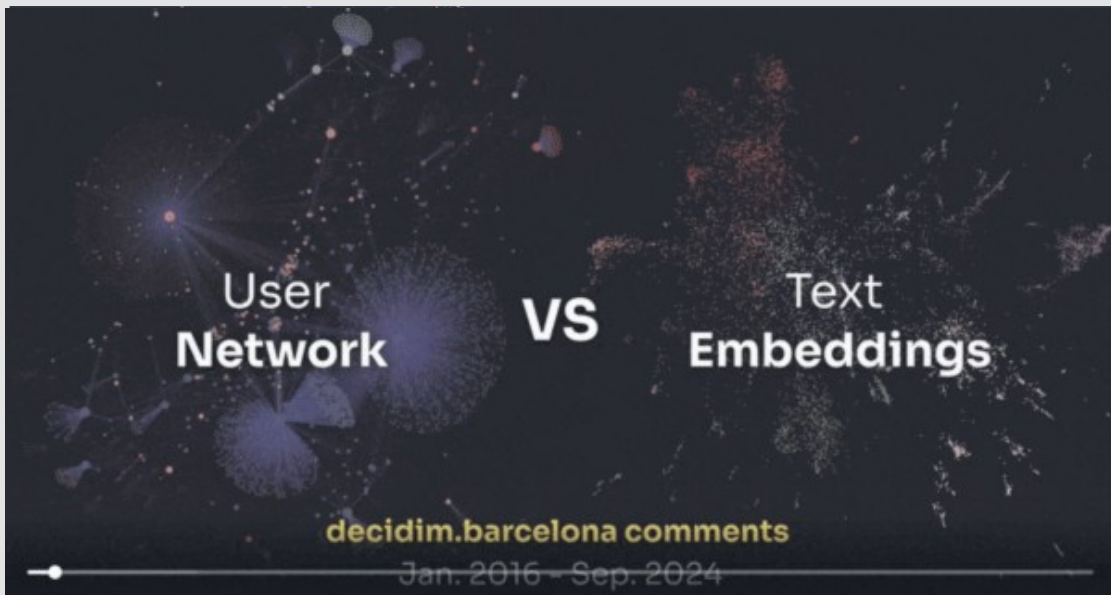
Python widget for interactive visualization in Jupyter Notebooks



GPL-3.0-or-later License

User networks VS embeddings

31,899 comments from Decidim Barcelona 2016-2024



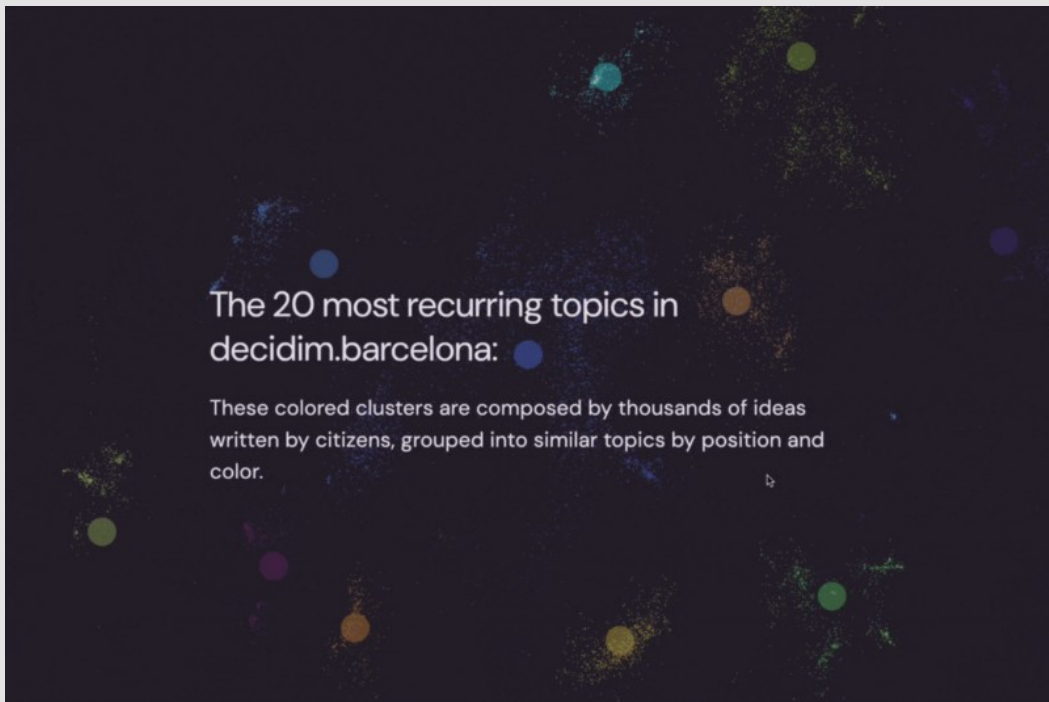
User2process network

Semantic embeddings

- **Cosmograph** comparison of user2process network VS 2D semantic embeddings, clustered with UMAP and K-means.
- Dynamic visualization from **31,899 comments** from 8,161 users, from 2016 to 2024 across 200 participatory processes.
- Project AINA **catalan fine-tuned sentence embeddings model** (*ST-NLI-ca_paraphrase-multilingual-mpnet-base*), by Barcelona Supercomputing Center (BSC).

Top 20 topics from decidim proposals

31,775 proposals from Decidim Barcelona 2016-2024



- **Cosmos.gl “scrollytelling”**
custom web interface, alternating network and 2D embeddings views, with Scrollama and Chart.js.
- **Top 20 topics modelled from 31,775 proposals** with BERTopic (59.4% with valid topics).
- Project AINA **catalan fine-tuned sentence embeddings model** (*ST-NLI-ca_paraphrase-multilingual-mpnet-base*), by Barcelona Supercomputing Center (BSC).

Top 20 topics from decidim proposals

34,792 proposals from **Brasil Participativo** 2023 -2025



- **Cosmos.gl “scrollytelling”**
custom web interface,
alternating network and 2D
embeddings views, with
Scrollama and Chart.js.
- **Top 20 topics modelled from
34,792 proposals** with
BERTopic (49.4% with valid
topics).
- Brazilian **portuguese fine-tuned
sentence embeddings model**
(tgsc/sentence-transformer-ult5-
pt-small). *Currently testing
serafim-900m-portuguese-pt-
sentence-encoder

 **Examining the margins,
limitations and biases in
embedding representations.**

 **Biases and limitations of pretrained models**, by their data and training design choices.



- ▶ **Proactively seek, flag, and mitigate biases** and risks (with human validation).
- 📋 Choose the **most adequate model** for the task and its scope.
- 📖 **Prepare relevant and diverse context** as knowledge base.
 - **At inference (RAG)**, achievable with less resources.
 - **Fine-tuning*** can help, but it needs bigger data and compute resources (*DPI initiatives could help).

 Training and inference **economic and environmental costs**.



- 🍷 **Less resource-intensive, accessible and context-specific** models. Prioritize what is needed for grounded, relevant and inclusive representation of knowledge.
- 🗑️ **Give up unnecessary performance and features** (“real-time”, unlimited generation, generalization to “any” context).

🔊 (Estimated) semantic preponderance **IS NOT equivalent to consensus**



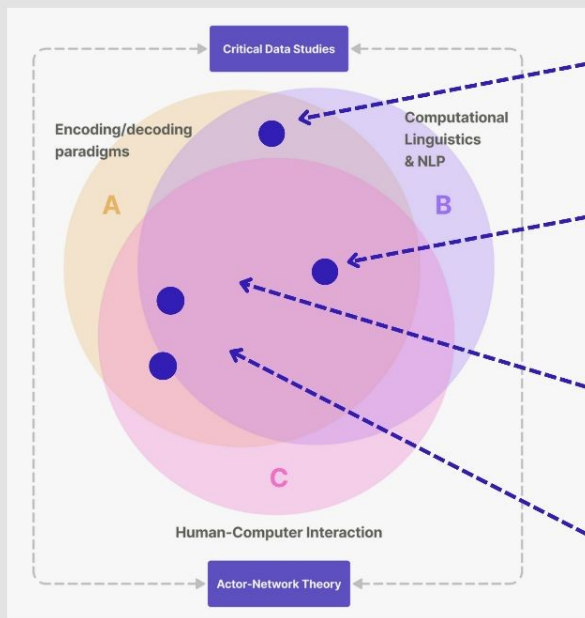
- 👏 Incorporate facilitation and human feedback for **guided topic modelling**, which integrate deliberative and diverse inputs.
- 🧠 **Connect high-dimensional spaces**, such as semantic space (texts) with to the agreement space (votes).

**Beyond human alignment,
distributed agency
and sovereignty
over our systems 🙌**

Part of a larger personal research project

***Emergent affordances for scaling qualitative readings
of large collections of multimodal texts***

(MA Final Project / PhD Proposal)



Chen et. al., 2023

*What We Can Do and Cannot Do with
Topic Modeling: A Systematic Review.*

Huang et. al. (2023)

*VA + Embeddings STAR: A State-of-the-
Art Report on the Use of Embeddings in
Visual Analytics.*

Arnold, T., & Tilton, L. (2019)

*Distant viewing: Analyzing large
visual corpora.*

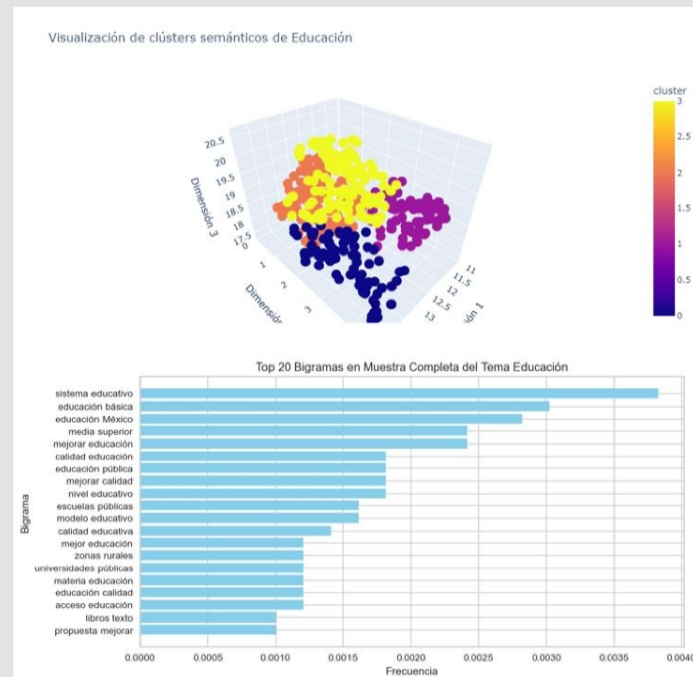
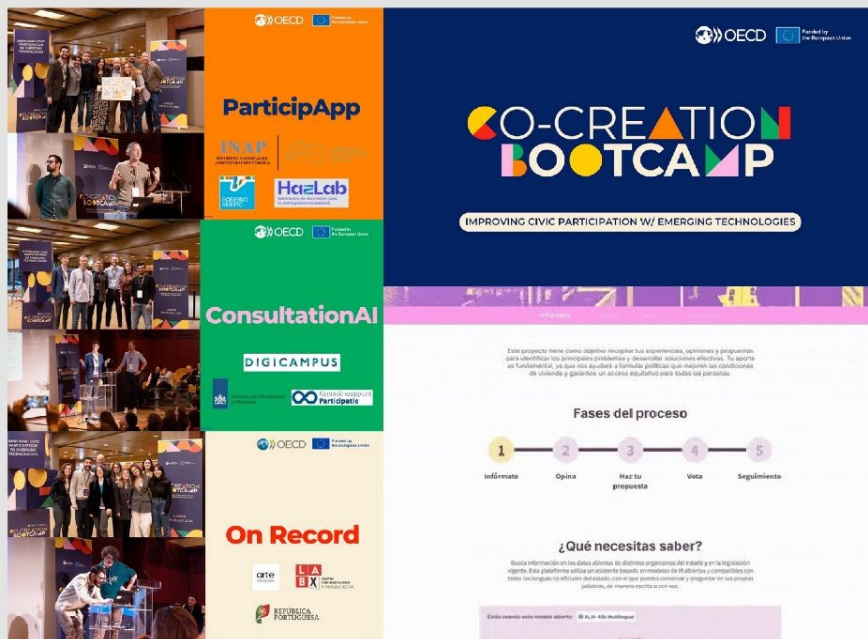
Oiva et al. (2024)

Soviet View of the World.



**Universitat
Pompeu Fabra
Barcelona**

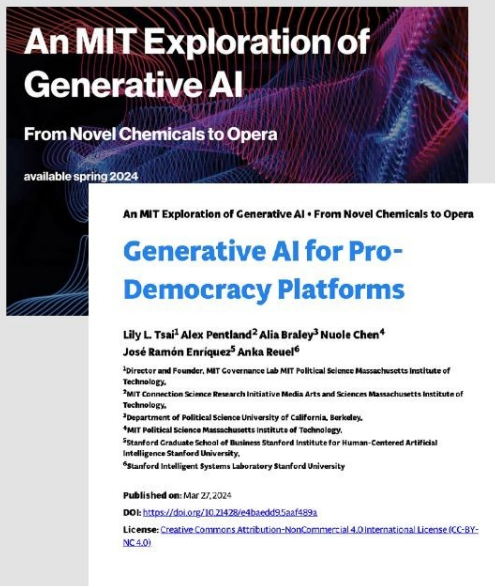
Related projects and collaborations



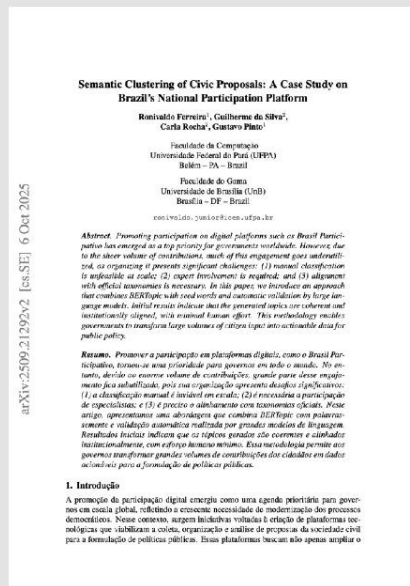
Improving Civic Participation with Emerging Technologies
(OECD-OPSI, 2025) Co-Creation Bootcamp and prototypes

Preguntas DebateINE (Signa Lab ITESO, 2024)
Question selection for Mexican presidential debate

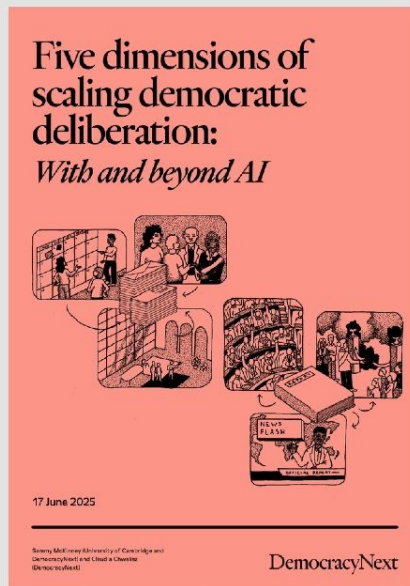
Other relevant research



Tsai, L. L., Pentland, A., Braley, A., Chen, N., Enríquez, J. R., & Reuel, A. (2024). *Generative AI for Pro-Democracy Platforms*. An MIT Exploration of Generative AI. MIT OpenPublishing



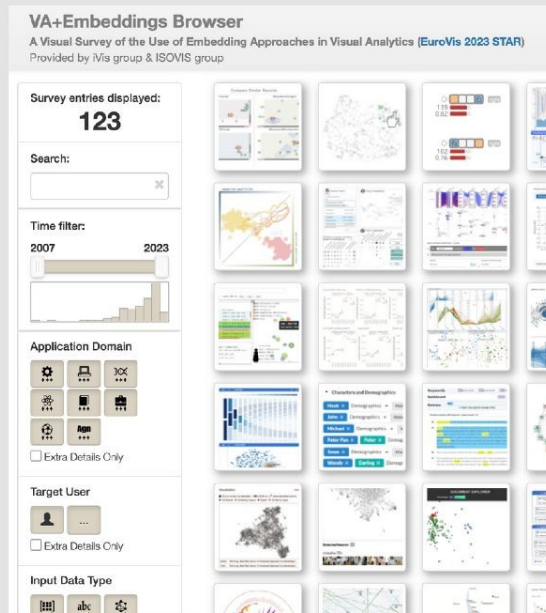
Ferreira, R., da Silva, G., Rocha, C., & Pinto, G. (2025). *Semantic Clustering of Civic Proposals: A Case Study on Brazil's National Participation Platform*.



McKinney, Sammy and Claudia Chwalisz (2025). *Five dimensions of scaling democratic deliberation: With and beyond AI*. DemocracyNext.



Chen, Y., Peng, Z., Kim, S. -H., & Choi, C. W. (2023). *What We Can Do and Cannot Do with TopicModeling: A Systematic Review.* Communication Methods and Measures



Huang et. al. (2023)
VA + Embeddings STAR: A State-of-the-Art Report on the Use of Embeddings in Visual Analytics. Computer Graphics Forum.



Small, C. T., Vendrov, I., Durmus, E., Homaei, H., Barry, E., Cornebise, J., ... & McGill, C. (2023). *Opportunities and risks of LLMs for scalable deliberation with Polis.*

References:

Arnold, T., & Tilton, L. (2023). **Distant viewing: Computational exploration of digital images**. The MIT Press. <https://doi.org/10.7551/mitpress/14046.001.0001>

Chen, Y., Peng, Z., Kim, S. -H., & Choi, C. W. (2023). **What We Can Do and Cannot Do with Topic Modeling: A Systematic Review**. *Communication Methods and Measures*, 17(2), 111-130. <https://doi.org/10.1080/19312458.2023.2167965>

Grootendorst, M. (2022). **BERTopic: Neural topic modeling with a class-based TF-IDF procedure**. arXiv preprint arXiv:2203.05794

Ferreira, R., da Silva, G., Rocha, C., & Pinto, G. (2025). **Semantic Clustering of Civic Proposals: A Case Study on Brazil's National Participation Platform**. arXiv preprint arXiv:2509.21292.

Huang, Z., Witschard, D., Kucher, K., & Kerren, A. (2023). **VA + Embeddings STAR: A State-of-the-Art Report on the Use of Embeddings in Visual Analytics**. *Computer Graphics Forum*, 42(3), 539-571. <https://doi.org/10.1111/cgf.14859>

McInnes, L., Healy, J., & Melville, J. (2018). **Umap: Uniform manifold approximation and projection for dimension reduction**. arXiv preprint arXiv:1802.03426.

McKinney, Sammy and Claudia Chwalisz (2025). **Five dimensions of scaling democratic deliberation: With and beyond AI, DemocracyNext**. <https://www.demnext.org/projects/five-dimensions-of-scaling-democratic-deliberation-with-and-beyond-ai>

Mendes, F.M., Poppi, R., Parra, H., Moreira, B. (2019). **EJ: A Free Software Platform for Social Participation**. In: Bordeleau, F., Sillitti, A., Meirelles, P., Lenarduzzi, V. (eds) *Open Source Systems*. OSS 2019. IFIP Advances in Information and Communication Technology, vol 556. Springer, Cham. <https://doi.org/10.1007/978-3-030-20883-7>

Mikolov, T. (2013). **Efficient estimation of word representations in vector space**. arXiv preprint arXiv:1301.3781, 3781.

OECD (2025). **Tackling civic participation challenges with emerging technologies**, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/2cb37a30-en>.

OECD (2025). **How Innovation Ecosystems Foster Citizen Participation Using Emerging Technologies in Portugal, Spain and the Netherlands**, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/2cb37a30-en>.

Reimers, N., & Gurevych, I. (2019). **Sentence-bert: Sentence embeddings using siamese bert-networks**. arXiv preprint arXiv:1908.10084.

Small, C. T., Vendrov, I., Durmus, E., Homaei, H., Barry, E., Cornebise, J., ... & McGill, C. (2023). **Opportunities and risks of LLMs for scalable deliberation with Polis**. arXiv preprint arXiv:2306.11932.

Signal_Lab ITESO. (2024). **preguntas-debateINE-2024**. GitHub. <https://github.com/signalab/preguntas-debateINE-2024>

Tsai, L. L., Pentland, A., Braley, A., Chen, N., Enríquez, J. R., & Reuel, A. (2024). **Generative AI for Pro-Democracy Platforms**. An MIT Exploration of Generative AI. <https://doi.org/10.21428/e4baedd9.5aaf489a>

THANK YOU!

DecidimFest 2025

Lightning talks - Round 2: Artificial Intelligence

Diego Arredondo Ortiz

